Genetic Epidemiology 17:157–164 (1999)

Regional Inference With Averaged *P* Values Increases the Power to Detect Linkage

Lynn R. Goldin,[1]* Gary A. Chase,[2] and Alexander F. Wilson[3]

[1]*Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, Maryland*
[2]*Department of Biostatistics and Research Epidemiology, Henry Ford Health System, Detroit, Michigan*
[3]*Genometrics Section, Inherited Disease Research Branch, National Human Genome Research Institute, NIH, Baltimore, Maryland*

Controversy exists with respect to the choice of an appropriate critical value when testing for linkage in a genomic screen. A number of critical values have been proposed for single-locus and multi-locus linkage analyses. In this study, criteria based on multiple single-locus analyses (i.e., regional test criteria) are evaluated using simulation methods for three different map densities. Tests based on single loci, multiple consecutive single loci, and moving averages of consecutive single loci are considered. Appropriate critical values are determined based on results from simulations under the null hypothesis of no linkage. The power of each "regional test " was compared to the power of a single-locus test. Results suggest that the best power was found when averaging *P* values over an interval size of 9–15 cM, and that testing the average of *P* values from two consecutive loci is superior to testing each single locus separately. The increase in power ranged from 7– 29% over the simulations considered. Genet. Epidemiol. 17:157–164, 1999. © 1999 Wiley-Liss, Inc.

Key words: regional inference; linkage detection; genome screening; sib-pair tests

*Correspondence to: Lynn R. Goldin, Genetic Epidemiology Branch, DCEG, NCI, 6120 Executive Blvd., MSC 7236, Bethesda, MD 20892-7236. E-mail: goldinl@exchange.nih.gov

Contract grant sponsor: National Center for Research Resources; Contract grant number: 1P41 RR03655.

Received for publication 9 April 1998; revision accepted 1 December 1998

INTRODUCTION

Genome mapping studies are being carried out on a number of complex diseases where the susceptibility genes involved are likely to have moderate to small effects, requiring investigation of large numbers of families [e.g., see Hauser et al., 1996]. It has also been suggested that when an entire genome is scanned for linkage, the nominal *P* value for "significance" at a specific location should be quite small to allow for the large number of tests. Lander and Kruglyak [1995] have suggested that *P* values on the order of 10^{-5} are needed to conclude that there is "significant" evidence for linkage; the test statistic is assumed to have a posterior probability of 5% for a false positive finding. However, in linkage studies of real and simulated data [e.g., Berrettini et al. 1997; Goldin et al., 1995], instead of seeing a highly significant *P* value at a single locus, one often sees "moderately significant" *P* values across a number of markers in a small region. Criteria for the detection of linkage based on regional behavior of the *P* value have been proposed [Goldin and Chase, 1997]. Such methods are given additional credibility by findings that around a true peak, the linkage test statistic may behave differently than in the vicinity of a false positive peak [Terwilliger et al., 1997]. In that study, it was shown (both by theoretical arguments and by simulation) that true linkage peaks in genome scans are wider than false peaks, and it was suggested that observing the length of linkage peaks was most useful for eliminating false positive peaks. Siegmund et al. [1997] made a similar observation in a single simulated sample from the Genetic Analysis Workshop 10 [MacCluer et al., 1997] using data simulated for several quantitative trait measurements and marker genotypes. They computed the average IBD value over a series of markers and found that many false positive findings were eliminated this way. Using the same series of simulated data from Genetic Analysis Workshop 10, Goldin and Chase [1997] analyzed 100 replicate samples and found that linkage criteria based on a series of *P* values appeared to be more powerful than criteria based on single locus tests.

In the current study, the power of alternative criteria for linkage is examined more systematically. Relatively simple genetic models were simulated and the power to detect linkage of several multiple point *P* value criteria was compared to the power to detect linkage using single point *P* value criteria, holding the false positive rate the same for all of the tests.

METHODS

A quantitative trait determined by a single two-allele locus was simulated in samples of 100 nuclear families consisting of two parents and two offspring with trait and marker information known for all individuals (100 sibpairs). The heritability of the single locus component was either 90, 75, or 50%, with the remainder of the phenotypic variation due to random environmental effects. A map of 11 marker loci (10 map intervals) was simulated, with each marker having 4 equally frequent alleles (heterozygosity of 0.75) . The trait locus was located between markers six and seven. Three different map densities were used. Markers were equally spaced at densities of 3, 5, and 10 cM, with the trait locus set at 16.5, 27.5, and 55 cM, respectively. The simulated chromosome segments were 30, 50, and 100 cM in length. Not

all combinations of parameters were simulated. Ascertainment of families was independent of trait values.

The Genometric Analysis Simulation Program (GASP) [Wilson et al., 1996] was used to simulate the samples. One thousand samples were simulated for each power determination. Each sample contained a single chromosomal segment, with the locus responsible for the trait located in the middle of the segment as described above. For each marker in each sample, the Haseman-Elston [Haseman and Elston, 1972] test statistic was computed using the SIBPAL program v2.7 [SAGE, 1994]. The resulting statistics and *P* values were analyzed using the SAS program.

Several different linkage tests involving *P* values at more than one locus were examined in order to compare the power of these regional inference rules to the power of a single locus test. Thresholds for both "significant" and "suggestive" linkage as defined by Lander and Kruglyak [1995] were used. A significant linkage would occur with probability no greater than 0.05 for a single genome scan under the null hypothesis; a suggestive linkage would occur once per genome scan.

In order to set the appropriate thresholds for the different inference rules, the same marker maps described above (with 11 markers and 10 map intervals) were simulated, but the trait locus was located on a different chromosome and not linked to any of the markers on the chromosome map segment. For each of the three map densities, 1,000 genome scans (each of 3,300 cM) were simulated. Thus, a single genome scan for the 10-cM map contained 33 100-cM chromosomal segments, a single genome scan for the 5-cM map contained 66 50-cM chromosomal segments, and the 3 cM map contained 110 30-cM chromosomal segments. The thresholds were set so that significant linkage would occur approximately 50 times in 1,000 scans and suggestive linkage would occur approximately 1,000 times in 1,000 genome scans (once/scan). The following multiple *P* value criteria were tested: 2 consecutive *P* values; 2 of 3 consecutive *P* values; and the average of 2, 3, or 4 consecutive *P* values. For each of these criteria, overlapping computations were done by advancing the set of *P* values by one marker along the segment.

RESULTS

The thresholds for single and multiple *P* values for each criterion were determined from the simulations under the null hypothesis of no linkage as described above. Threshold values were tested by trial and error until the desired false positive rates were obtained. These thresholds are shown in Table I for each of the three map densities. Thresholds were set to keep the false positive rate identical (or nearly identical) for each criterion examined. The thresholds for single *P* values were larger than the 2.2×10^{-5} (significant) and 7.4×10^{-4} (suggestive) critical values suggested by Lander and Kruglyak [1995] for the case of a continuous genome scan for fully informative markers. As expected, the thresholds were less stringent for more widely spaced markers and the thresholds for all criteria became more stringent as the map density increased.

The comparison of power for the different criteria is shown in Table II for the trait with 90% heritability for all three map densities. No improvement in power or a very small improvement was seen in some cases for the criteria of 2 consecutive significant *P* values or 2 of 3 consecutive significant *P* values compared to a single *P*

TABLE I. Thresholds for Single Point *P* Values and Multiple *P* Values*

Criterion		Threshold for: 10 cM Map	5 cM Map	3 cM Map
Single *P* value	Significant	0.00018	0.00012	0.000095
	Suggestive	0.00332	0.0019	0.001175
2 consecutive *P* values	Significant	0.00305	0.0012	0.00083
	Suggestive	0.02314	0.01275	0.00775
2/3 consecutive *P* values	Significant	0.0025	0.0011	0.00065
	Suggestive	0.01947	0.0104	0.00615
Average of 2 *P* values	Significant	0.00205	0.00092	0.0006
	Suggestive	0.01622	0.00895	0.00546
Average of 3 *P* values	Significant	0.0056	0.0027	0.00158
	Suggestive	0.03525	0.0183	0.01132
Average of 4 *P* values	Significant	0.0125	0.0055	0.0028
	Suggestive	0.05695	0.0295	0.01755

*All thresholds set so that the number of false positives was 50 (±1) for significant linkage and 1,000 (±2) for suggestive linkage.

value criterion. However, consistent improvement (up to 10 percentage points) was seen using average *P* values. In the case of the 10-cM map, the largest increase in power was seen when 2 consecutive *P* values were averaged (10-cM interval). This was expected since linkage evidence decreases as the distance between the marker and trait locus increases. For the 5- and 3-cM maps, an average of 4 *P* values (interval size of 15 and 9 cM, respectively) gave the highest power. An average of 5 *P* values (12-cM interval) on the 3-cM map showed a slight additional increase in power for suggestive linkage (data not shown). It is interesting to note in Table II, the power improves when the map density increases from 10 to 5 cM but for the threshold of suggestive linkage, power *decreases* slightly when the density increases from 5 to 3 cM. In this case, the decrease is not substantial given the variation expected from the simulations. However, this does demonstrate that under the model and sample size used here, a "ceiling" in power is reached at a 5-cM map density and that there is not enough information gained by further increasing the marker density to outweigh the more stringent *P* value threshold required.

TABLE II. Power of Linkage Detection for a Quantitative Trait With 90% Heritability

Criterion	Linkage threshold	Power (%) for 90% heritability: 10 cM	5 cM	3 cM
Single *P* value	Significant	31.5	39.0	45.6
	Suggestive	76.4	80.2	79.9
2 consecutive *P* values	Significant	38.1	39.9	46.7
	Suggestive	77.1	83.3	82.2
2/3 consecutive *P* values	Significant	38.0	43.5	49.7
	Suggestive	79.1	84.8	83.8
Average of 2 *P* values	Significant	38.8	44.2	47.6
	Suggestive	79.7	86.0	83.6
Average of 3 *P* values	Significant	36.0	47.6	51.2
	Suggestive	80.0	87.1	85.4
Average of 4 *P* values	Significant	34.2	49.8	52.1
	Suggestive	78.9	88.0	86.3

Table III shows analogous results for a trait with 75% heritability. Again, there is a consistent increase in power when averaging *P* values in a region compared to testing single *P* values. For the 3-cM map, we also tested a trait with 50% heritability. The power to detect linkage was uniformly low in this case due to the relatively few sibpairs and comparisons are not shown. However, because traits with low heritability are of interest for mapping complex traits, some additional simulations were performed for traits with 50 and 30% heritability but using larger sample sizes. Table IV shows the results for these models assuming a 5-cM map density. For simplicity, only the power comparisons between a single *P* value and an average of 4 consecutive *P* values are shown. It is evident that average *P* values in a region are also more powerful when heritability is lower.

Table V shows the percentage improvement according to heritabiliity averaged over the three map densities. The improvement for 75% heritability is slightly larger than that for 90% heritability. The absolute power for significant linkage is lower than that for suggestive linkage, and thus the percentage improvement is expected to be larger.

The degree of improvement according to the position of the trait locus relative to the marker map is also considered. Table VI shows the power of the different criteria for the case of 90% heritability and a 5-cM map. As expected, when the trait locus is outside the map (in this case, 2.5 cM telomeric from marker locus 1), the power to detect linkage decreases when multiple *P* values are used. When the trait locus is between the first two markers, the power is the same for single *P* values and multiple *P* values although power starts to decrease when 3 or more *P* values are averaged. When the trait is between markers 2 and 3, there is an improvement in power but not as much as when the trait locus is in the middle of the map. These calculations with the trait locus beyond or at one end of the map are relevant only to the extent that there are gaps in the current human genome map or if the trait locus is at the end of a chromosome.

DISCUSSION

This study demonstrated that inference rules based on the average of a few *P* values in a small region improves the power to detect linkage. In one respect, it can be consid-

TABLE III. Power of Linkage Detection for a Quantitative Trait With 75% Heritability

		Power (%) for 75% heritability		
Criterion	Linkage threshold	10 cM	5 cM	3 cM
Single *P* value	Significant	9.6	13.0	15.5
	Suggestive	43.5	45.7	46.5
2 consecutive *P* values	Significant	13.0	12.6	14.6
	Suggestive	48.7	48.6	50.3
2/3 consecutive *P* values	Significant	12.6	14.2	16.1
	Suggestive	48.9	51.5	50.4
Average of 2 *P* values	Significant	12.8	14.1	16.0
	Suggestive	49.3	49.7	50.5
Average of 3 *P* values	Significant	11.7	15.8	17.9
	Suggestive	50.1	54.4	54.9
Average of 4 *P* values	Significant	11.7	15.8	17.9
	Suggestive	51.7	56.5	56.7

TABLE IV. Power of Linkage Detection for 50% and 30% Heritability

		Power (%) for heritability	
Criterion	Linkage threshold	50% (n = 300)	30% (n = 500)
Single *P* value	Significant	13.4	4.1
	Suggestive	47.1	21.0
Average of 4 *P* values	Significant	18.6	5.0
	Suggestive	53.4	27.3

ered a "smoothing" technique since marker information at single points will vary. However, even as the marker map got denser, there were improvements in power consistent with true linkage peaks being longer than false peaks. This method will identify linkage peaks that are present for several markers although the optimum window for testing depends on map density and marker informativeness. In the simulations considered, the best power was found when averaging statistics for an interval size of 9–15 cM. It was encouraging that even for 10-cM maps, testing the average of 2 consecutive *P* values was superior to testing each single *P* value. The very upper or lower limits of map density where averaging is useful was not determined in this study, but the method is clearly valuable for the range of densities that are commonly used in genome scans. We did not consider the effect of marker heterozygosity in this study. Clearly, a lower heterozygosity would decrease the absolute power of linkage detection but one would still expect averaged *P* values to be more powerful than single point *P* values. The marker heterozygosity used is comparable to that of current microsatellite markers; it is fair to say that even with current genotyping technology, marker informativeness is not a practical limitation of genetic mapping.

The question remains as to what should the exact thresholds for significant or suggestive linkage be when average *P* values are used? To what extent are they model dependent? There is no one criteria that can be applied to all cases. As in the single point case, the required *P* values for averages did depend on map density. If one examines the thresholds in Table I, it can be seen that they are all fairly close, even for different map densities. For a test interval of approximately 10–15 cM, the thresholds are between 0.002 and 0.005, and between 0.02 and 0.03 for significant and suggestive linkages, respectively.

As stated above, calculating an average of two or more single *P* values can be considered a smoothing technique. One can argue that multipoint calculations are a better smoothing technique since they make optimal use of linkage information. However, even in standard multipoint calculations, *P* values or lod scores are examined at single points along the genome; thus one would still expect to see an improvement using averages for several points in a region. This would be similar to the mulitpoint IBD method of Goldgar [1990], which tests for increased IBD sharing over regions rather than single points. In

TABLE V. Percentage Improvement of Power for Average *P* Values Combined Over Map Densities

Linkage threshold	90% Heritability	75% Heritability
Significant	22	29
Suggestive	7	20

TABLE VI. Effect of Trait Locus Location on Power of Multiple *P* Values

		Power of linkage detection : 90% heritability, 5-cM map		
Criterion	Linkage threshold	Trait locus outside map	Trait locus between markers 1 and 2	Trait locus between markers 2 and 3
Single *P* value	Significant	26.8	35.8	38.9
	Suggestive	65.1	74.9	79.5
2 consecutive *P* values	Significant	19.0	34.3	41.0
	Suggestive	58.7	75.0	81.1
2/3 consecutive *P* values	Significant	21.6	36.6	44.3
	Suggestive	58.6	75.0	82.7
Average of 2 *P* values	Significant	21.0	36.6	43.9
	Suggestive	61.8	76.4	82.5
Average of 3 *P* values	Significant	17.7	30.7	45.0
	Suggestive	53.1	70.1	84.0
Average of 4 *P* Values	Significant	12.9	26.6	43.1
	Suggestive	46.6	63.7	80.3

addition, this regional test method should be relatively robust with respect to errors in marker order over small intevals. As map density increases, the increase in computational time for multiple average *P* values will be trivial, whereas the compuational time required for multipoint methods will increase tremendously over similarly sized mapping intervals. Finally, multiple average *P* values can be applied to any two point linkage statistic (e.g., sib-pair or variance components methods).

Our strategy in examining inference rules that utilize multiple *P* values could be expanded to consider the behavior of linkage test statistics as a nearly continuous processes. Using this approach, a curvilinear fit could be constructed for closely spaced markers and its behavior studied under null and alternative hypotheses. Then, a measure of discrepancy from the null hypothesis such as a mean integrated square error could be used as a test statistic for linkage.

Although full coverage of this topic is beyond the scope of the present work, it offers an interesting alternative for future investigation.

In summary, we showed that testing for linkage by computing average *P* values in a region has somewhat higher power than testing single point *P* values. The exact thresholds depend on map density but are all within a narrow range and are well suited to initial genome screens.

ACKNOWLEDGMENTS

We thank Alexa J. Sorant for her help in installing the GASP program. The results of this paper were obtained by using the program package SAGE, which is supported by a US. Public Health Service Resource Grant (1P41 RR03655) from the National Center for Research Resources.

REFERENCES

Berrettini WH, Ferraro TN, Goldin LR, Detera-Wadleigh SD, Choi H, Guroff JJ, Kazuba DM, Nurnberger JI Jr, Hsieh WT, Hoehe MR, Gershon ES. 1997. A linkage study of bipolar illness. Arch Gen Psychiatry 54:27–35.

Goldgar DE. 1990. Multipoint analysis of human quantitative genetic variation. Am J Hum Genet 47:957–967.

Goldin LR, Chase GA. 1997. Improvement of the power to detect complex disease genes by regional inference procedures. Genet Epidemiol 14:785–789.

Goldin LR, Chase GA, King TM, Badner JA, Gershon ES.1995. Use of exact and adjusted liability scores to detect genes affecting common traits. Genet Epidemiol 12:765–769.

Haseman JK, Elston RC. 1972.. The investigation of linkage between a quantitative trait and a marker locus. Behav Genet 2:3–19.

Hauser ER, Boehnke M, Guo S-W, Risch N. 1996. Affected-sib-pair interval mapping and exclusion for complex genetic traits: sampling considerations. Genet Epidemiol 13:117–138.

Lander E, Kruglyak L. 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nature Genet 11:241–247.

MacCluer JW, Blangero J, Dyer TD, Speer MC. 1997. GAW10: Simulated family data for a common oligogenic disease with quantitative risk factors. Genet Epidemiol 14:737–742.

SAGE. 1994. Statistical analysis for genetic epidemiology, release 2.2. Computer program package available from the Department of Epidemiology and Biostatistics. Cleveland: Case Western Reserve University.

Siegmund KD, Todorov AA, Rao DC, Borecki IB. 1997. The role of smoothing techniques in the interpretation of results from genomic scans using sib-pair data. Genet Epidemiol 14:1047–1052.

Terwilliger JD, Shannon WD, Lathrop GM, Nolan JP, Goldin LR, Chase GA, Weeks D. 1997. True and false positive peaks in genome-wide scans: applications of length-biased sampling to linkage mapping. Am J Hum Genet 61:430–438.

Wilson AF, Bailey-Wilson JE, Pugh EW, Sorant AJM. 1996. The genometric analysis simulation program (GASP): a software tool for testing and investigating methods in statistical genetics. Am J Hum Genet 59:A193.